The Importance of Tie-Breaking in Finite-Blocklength Bounds

Eli Haim
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: elih@eng.tau.ac.il

Yuval Kochman
School of CSE, HUJI
Jerusalem, Israel
Email: yuvalko@cs.huji.ac.il

Uri Erez
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: uri@eng.tau.ac.il

***Abstract*—We consider upper bounds on the error probability in channel coding. We derive an improved maximum-likelihood union bound, which takes into account events where the likelihood of the correct codeword is tied with that of some competitors. We compare this bound to various previous results, both qualitatively and quantitatively. With respect to maximal error probability of linear codes, we observe that when the channel is additive, the derivation of bounds, as well as the assumptions on the admissible encoder and decoder, simplify considerably.**

I. Introduction

Consider maximum-likelihood decoding, known to be optimal in the sense of average error probability between equiprobable messages. What happens when ℓ false codewords share the maximum likelihood score with the transmitted one? No matter how such a tie is broken, the average error probability given this event is $1 - 1/\ell+1$. Computing the optimal error probability, taking into account all possible ties, is exponentially hard. Can we ignore this event, i.e., assume that in case of a tie the decoder is always right or always wrong? The answer depends upon both the channel and the blocklength. When the likelihood score is a continuous random variable, the probability of ties is zero. Also, for long enough blocks, the distribution of the score of a word can be closely approximated by a continuous one (e.g., using the central-limit theorem). However, for small enough alphabet size and short enough blocks, the effect of ties on error-probability bounds is not negligible.

We revisit the finite-blocklength achievability results of Polyanskiy et al. [1]. For i.i.d. codewords, and when we can neglect ties, computation of the exact average error probability is not harder than that of the random-coding union (RCU) achievability bound. However, ties cannot always be neglected. As the RCU bound assumes that ties always lead to errors, it can be improved; indeed, we derive a tighter bound. In particular, unlike the RCU bound, the new bound is always tighter than bounds based upon threshold decoding.

When it comes to maximal error probability, tie-breaking is no longer a mere issue of analysis. Rather, ties have to be broken in a manner that is "fair", such that the error probability given different messages is balanced. In [1], a randomized decoder is employed in order to facilitate such fairness. But is randomization necessary? We show that at least for additive channels (over a finite field), a deterministic decoder suffices.

II. Notation and Background

We consider coding over a memoryless channel with some finite blocklength n, i.e.:

$$V(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} V(y_i|x_i). \tag{1}$$

for every $\mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n$. The channel input and output alphabets are arbitrary. For the sake of simplicity, we adopt discrete notation; the bounds do not depend on alphabet sizes, and the educated reader can easily translate the results to the continuous case (which is of limited interest in the context of tie-breaking). The codebook is given by $\mathbf{x}_1, \ldots, \mathbf{x}_M$, where M is the number of codewords (that is, the coding rate is $R = 1/n \log M$). The decoder produces an estimate $\hat{m}$, where the transmitted message index is denoted by m. The average error probability, assuming equiprobable messages, is given by:

$$\epsilon = \frac{1}{M} \sum_{m=1}^{M} \mathbb{P}\left(\hat{m} \neq m | \mathbf{X} = \mathbf{x}_m\right). \tag{2}$$

The maximum error probability is given by

$$\overline{\epsilon} = \max_{m=1\ldots M} \mathbb{P}\left(\hat{m} \neq m | \mathbf{X} = \mathbf{x}_m\right). \tag{3}$$

For the sake of analyzing the error probability, it is convenient to consider code ensembles. All ensembles we consider in this work fall in the following category.

Definition 1 (Random conditionally-symmetric ensemble): An ensemble is called random conditionally-symmetric ensemble (RCSE) if its codewords are drawn such that for every different $m, j, k \in \{1, \ldots M\}$ and for every $\mathbf{x}, \bar{\mathbf{x}} \in \mathcal{X}^n$:

$$\mathbb{P}\left(\mathbf{X}_j = \bar{\mathbf{x}} | \mathbf{X}_m = \mathbf{x}\right) = \mathbb{P}\left(\mathbf{X}_k = \bar{\mathbf{x}} | \mathbf{X}_m = \mathbf{x}\right) \tag{4}$$

It is easy to verify, that for an RCSE, all words are identically distributed. We can thus define by $\mathbf{X}$ a word drawn under the ensemble distribution (not necessarily memoryless) P over the set $\mathcal{X}^n$. Using this input distribution, the information density is given by:

$$i(\mathbf{x}; \mathbf{y}) = \log \frac{V(\mathbf{y}|\mathbf{x})}{PV(\mathbf{y})}, \tag{5}$$

where $PV(\mathbf{y})$ is the output distribution induced by $P(\mathbf{x})$ and

$V(y|x)$. We denote by $\mathbf{Y}$ the output corresponding to the random input $\mathbf{X}$, and the random variable $i(\mathbf{X};\mathbf{Y})$ is defined accordingly. In addition, we define $i(\bar{\mathbf{X}};\mathbf{Y})$ as the information density a codeword $\bar{\mathbf{X}}$ other[1] than the one that generated $\mathbf{Y}$.[2]

The importance of deriving bounds for an RCSE is due to the fact that this class includes many interesting ensembles. An important special case of RCSE is the pairwise-independent ensemble:

Definition 2 (Pairwise-independent ensemble): A pairwise independent ensemble (PIE) is an ensemble such that its codewords are pairwise-independent and identically distributed. That is, for any two indices $i \neq j$,

$$\mathbb{P}\left(\mathbf{X}_i = \mathbf{x}_i | \mathbf{X}_j = \mathbf{x}_j\right) = \mathbb{P}\left(\mathbf{X}_i = \mathbf{x}_i\right) = P(\mathbf{x}). \tag{6}$$

We note that the codewords of an RCSE are not necessarily pairwise-independent. One example is linear random codes with a cyclic generating matrix [2]. In this ensemble, a codebook is a linear code, such that all the cyclic shifts of the any codeword are also codewords. Generally, RCSE (which are not necessarily PIE) can be constructed by first drawing a class of codewords, and then, randomly (uniformly) drawing the codewords from this class. Alternatively, it can be constructed by choosing some codeword which defines the class, from which all the other codewords will be drawn.

Finally, the following class of channels turns out to play a special role.

Definition 3 (Additive channels): A channel is additive over a finite group $\mathcal{G}$ with an operation, if $\mathcal{X} = \mathcal{Y} = \mathcal{G}$, and the transition distribution $V(y|x)$ is compatible with

$$Y = X + N$$

where N is statistically independent of X, and "$+$" denotes the operation over $\mathcal{G}$.[3]

For example, for modulo-additive channels the alphabet is the ring $\mathbb{Z}_q$, and addition is modulo q. The importance of additive channels stems from the following.

Lemma 1: Consider an additive channel over $\mathcal{G}$, and a codebook drawn from a PIE with uniform input distribution over $\mathcal{G}^n$, i.e. $P(\mathbf{x}) = |\mathcal{G}|^{-n} \quad \forall \mathbf{x} \in \mathcal{G}^n$. Then, $i(\bar{\mathbf{X}};\mathbf{Y})$ is statistically independent of $(\mathbf{X},\mathbf{Y})$.

Proof: For this channel the information density (5) is equal to

$$i(\mathbf{x};\mathbf{y}) = \log \frac{P_{\mathbf{N}}(\mathbf{y}-\mathbf{x})}{P_{\mathbf{Y}}(\mathbf{y})}, \tag{7}$$

where $P_{\mathbf{N}}(\cdot)$ is the distribution of the noise, and $P_{\mathbf{Y}}(\cdot)$ is the distribution of the channel output. For this channel with codebook drawn from a PIE with a uniform distribution over $\mathcal{G}^n$, we have that for every $\mathbf{z} \in \mathcal{G}^n$:

$$\mathbb{P}\left(\mathbf{Y} - \bar{\mathbf{X}} = \mathbf{z}\right) = \mathbb{P}\left(\mathbf{X} + \mathbf{N} - \bar{\mathbf{X}} \ = \mathbf{z}\right) \tag{8a}$$

$$= |\mathcal{G}|^{-n}, \tag{8b}$$

since $\bar{\mathbf{X}}$ is uniformly distributed over $\mathcal{G}^n$ and statistically independent of $(\mathbf{X},\mathbf{N})$. Therefore, $\mathbf{Y}-\bar{\mathbf{X}}$ is statistically independent of $(\mathbf{X},\mathbf{Y})$; Moreover, any function of $\mathbf{Y}-\bar{\mathbf{X}}$ is also statistically independent of $(\mathbf{X},\mathbf{Y})$, in particular $P_N(\mathbf{Y}-\bar{\mathbf{X}})$ is statistically independent of $(\mathbf{X},\mathbf{Y})$.

Since $\mathbf{X}$ is uniformly distributed over $\mathcal{G}^n$, and is statistically independent noise, then the channel output $\mathbf{Y}$ is also uniformly distributed over $\mathcal{G}^n$, i.e. for any $\mathbf{y} \in \mathcal{G}^n$:

$$P_{\mathbf{Y}}(\mathbf{y}) = |\mathcal{G}|^{-n}, \tag{9}$$

and hence, $P_{\mathbf{Y}}(\mathbf{Y})$ is statistically independent of $(\mathbf{X},\mathbf{Y})$. From the two observations above, we conclude that $i(\bar{\mathbf{X}};\mathbf{Y})$ is statistically independent with $(\mathbf{X},\mathbf{Y})$. ■

III. I.I.D. Codebooks

Before stating the main results that apply to any RCSE, we start by simple bounds that hold for the special case of an i.i.d. ensemble. That is, all codewords are mutually independent, and each one distributed according to $P(\mathbf{X})$. In this case, the average error probability is well known, although hard to compute [1]. Denote:

$$W = \mathbb{P}\left(i(\bar{\mathbf{X}};\mathbf{Y}) = i(\mathbf{X};\mathbf{Y}) | \mathbf{X},\mathbf{Y}\right) \tag{10a}$$

$$Z = \mathbb{P}\left(i(\bar{\mathbf{X}};\mathbf{Y}) < i(\mathbf{X};\mathbf{Y}) | \mathbf{X},\mathbf{Y}\right). \tag{10b}$$

Then, for an i.i.d. ensemble [1, Thm. 15]:

$$\epsilon^{(\text{iid})} = 1 - \sum_{\ell=0}^{M-1} \frac{1}{\ell+1} \cdot \binom{M-1}{\ell} \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left(W^{\ell} Z^{M-1-\ell}\right). \tag{11}$$

This result stems from the fact that for equiprobable words, maximum likelihood (ML) decoding is just maximum information density. We note that ℓ represents the number of competing codewords that share the maximal information-density score with the correct one; given ℓ, the correct codeword will be chosen with probability $1/\ell+1$. If $W = 0$ (as happens when $V(Y|x)$ is a proper density for every $x \in \mathcal{X}$), the calculation is straightforward. Otherwise, it has exponential complexity. Thus, the main burden is with dealing with ties. In order to avoid such burden, we suggest the following simple bounds.

Proposition 1 (Bounds for i.i.d. codebooks): For an i.i.d. ensemble,

$$1 - \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[(W+Z)^{M-1}\right] \leq \epsilon^{(\text{iid})} \leq 1 - \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[Z^{M-1}\right]. \tag{12}$$

This result can be shown either from (11) or directly. For the lower bound, in case multiple codewords (including the correct one) attain the maximal information density, the correct one is always chosen; for the upper bound, it is never chosen under such circumstances. Of course, as the upper bound is just the first term in (11), one may tighten it by taking more terms.

[1] In a random codebook it may happen that the codebook contains some identical codewords. Thus it is possible that $\bar{\mathbf{X}} = \mathbf{X}$, as long as they represent different messages.

[2] In [1], the notation $i(\mathbf{X};\bar{\mathbf{Y}})$ is sometimes used; for RCSE, the two are equivalent.

[3] The operation "$-$" over the group, which is uniquely defined by the operation "$+$", such that for any $a,b,c \in \mathcal{G}$: $a - b = c$ iff $a = c + b$.

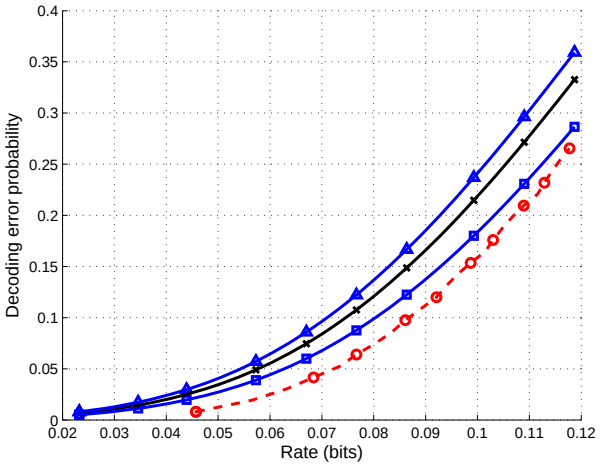

Fig. 1. The effect of tie-breaking on the performance of i.i.d. codebooks. We demonstrate the effect using a BSC with crossover probability 0.3, at blocklength $n = 100$. The triangle- and square- marked solid curves give the lower and upper bounds of Proposition 1, respectively. The $\times$-marked solid curve is the exact error probability of the i.i.d. ensemble (11), evaluated by taking enough terms in the sum, such that the effect of additional ones is numerically insignificant. For reference, the circle-marked dashed curve gives the tightest lower bound on the error probability, which holds for *any* codebook [1, Theorem 16].

The difference between the lower and upper bounds may be quite significant, as demonstrated in Figure 1.

IV. Bounds for RCSE

A. Maximum-Likelihood Union Bounds

When the codewords are not i.i.d., we cannot use anymore products of probabilities as done in the previous section. However, for providing a lower bound on the error probability, we can use a union bound. We derive a result that is close in spirit to the RCU bound [1, Theorem 16], which states that $\epsilon^{(\text{iid})} \leq \epsilon_{\text{RCU}}$,[4] where

$$\epsilon_{\text{RCU}} \triangleq \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot(1-Z)\right\}\right]. \tag{13}$$

We improve this bound in two ways: First, it is extended to any RCSE, and second, the case of equal maximal information-density scores is taken into account.

Theorem 1 (RCU* bound): The average error probability of an RCSE satisfies $\epsilon^{(\text{RCSE})} \leq \epsilon_{\text{RCU}^*}$ where

$$\epsilon_{\text{RCU}^*} \triangleq \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot\left(1-Z-\frac{W}{2}\right)\right\}\right], \tag{14}$$

where the conditional probabilities W and Z are given by (10).

Proof: Without loss of generality, assume that the transmitted codeword index is $m = 1$. The ML decoder will choose the codeword with maximal information density; in case of equality, it will uniformly draw a winner between the maximal ones. Let C_j be the event that the codeword j was chosen in such a lottery. Denote the following events:

$$A_j \triangleq \{i(\mathbf{X}_j;\mathbf{Y}) > i(\mathbf{X};\mathbf{Y})\} \tag{15a}$$

$$B_j \triangleq \{i(\mathbf{X}_j;\mathbf{Y}) = i(\mathbf{X};\mathbf{Y})\}. \tag{15b}$$

Also denote $A \triangleq \bigcup_{j=2}^{M} A_j$ and $B \triangleq \bigcup_{j=2}^{M} B_j$. Then, the error probability is given by:

$$\begin{aligned}\epsilon^{(RCSE)} &= \mathbb{P}\left(A \cup \left[B \cap \bar{C}_1\right] \middle| m=1\right) && \text{(16a)}\\ &= \mathbb{E}_{\mathbf{X}_1,\mathbf{Y}}\mathbb{P}\left(A \cup \left[B \cap \bar{C}_1\right] \middle| m=1,\mathbf{X}_1,\mathbf{Y}\right). && \text{(16b)}\\ &= \mathbb{E}_{\mathbf{X}_1,\mathbf{Y}} \min\left\{1, \mathbb{P}\left(A \cup \left[B \cap \bar{C}_1\right] \middle| m=1,\mathbf{X}_1,\mathbf{Y}\right)\right\} && \text{(16c)}\\ &\triangleq \mathbb{P}\left(A \cup \left[B \cap \bar{C}_1\right] \middle| m=1,\mathbf{X}_1,\mathbf{Y}\right). && \text{(16d)}\end{aligned}$$

Using the union bound between events of equality and inequality, we have:

$$S \leq \mathbb{P}\left(A | m=1,\mathbf{X}_1,\mathbf{Y}\right) + \mathbb{P}\left(B \cap \bar{C}_1 \middle| m=1,\mathbf{X}_1,\mathbf{Y}\right). \tag{17}$$

Now, the event C_1 depends on the the rest of the variables only through the number of codewords that achieve equal score. Specifically, if there are ℓ impostors, then $\mathbb{P}(C_1) = 1/\ell+1$. Since the second term is non-zero only if $\ell \geq 1$, it follows that:

$$S \leq \mathbb{P}\left(A | m=1,\mathbf{X}_1,\mathbf{Y}\right) + \frac{1}{2}\mathbb{P}\left(B | m=1,\mathbf{X}_1,\mathbf{Y}\right). \tag{18}$$

We now use the union bound, as in [1]:

$$S \leq \sum_{j=2}^{M} \mathbb{P}\left(A_j | m=1,\mathbf{X}_1,\mathbf{Y}\right) + \frac{1}{2}\sum_{j=2}^{M} \mathbb{P}\left(B_j | m=1,\mathbf{X}_1,\mathbf{Y}\right). \tag{19}$$

Noting that for an RCSE each element in the left (resp. right) sum equals $1 - W - Z$ (resp. W), and substituting this bound in (16d) we arrive at the desired result. ■

Remark 1: We can give the RCU* bound the following interpretation. First, each potential input $\mathbf{x}_j$ is given an information-density score (equivalent to a likelihood score) i_j. Then, these scores are fed to a comparison process. The process is biased against the correct codeword, in the sense that it has to beat each and every impostor. However, each pairwise comparison itself is optimal (the correct codeword will beat an impostor with lower score), and fair (in case of a tie, both codewords are equally likely to win). This comparison mechanism is worse than the actual decoder used in the proof, since in case the correct codeword shares the maximal score with ℓ impostors, it has probability $2^{-\ell}$ to be chosen, rather than $1/\ell+1$; yet, the union bound for both is equal.

B. Relation to Gallager's Type-I bound

The following bound is due to Gallager.

Proposition 2 (Gallager type-I bound [3], Sec. 3.3): For any constant t:

$$\epsilon^{(RCSE)} \leq \epsilon_{\text{G-I}}, \tag{20}$$

[4] Indeed, it is noted in [1, Appendix A] that pairwise independence is sufficient.

where

$$\epsilon_{\text{G-I}} \triangleq \mathbb{P}\left(i(\mathbf{X};\mathbf{Y}) < t\right) + \\ + (M-1)\mathbb{P}\left(i(\mathbf{X};\mathbf{Y}) \geq t \wedge i(\bar{\mathbf{X}};\mathbf{Y}) \geq i(\mathbf{X};\mathbf{Y})\right). \quad (21)$$

Just like the RCU, this bound is based upon a union bound for the ML decoder. However, it is inferior to the RCU bound, due to the following consideration. Taking the minimum between the union and one in the RCU bound is similar to the threshold t in (21), in the way that it avoids over-estimating the error probability in cases where the channel behavior was "bad". However, the RCU bound uses the optimal threshold given $\mathbf{X}$ and $\mathbf{Y}$; the Gallager bound uses a *global* threshold, which reflects a tradeoff. Nevertheless, for additive channels (recall Definition 3) the local and global thresholds coincide.

Proposition 3: For any RCSE:

$$\epsilon_{\text{G-I}} \geq \epsilon_{\text{RCU}}, \quad (22)$$

where $\epsilon_{\text{G-I}}$ and ϵ_{RCU} are defined in (21) and in (13) respectively. If the channel is additive and the code ensemble is PIE with uniform distribution over $\mathcal{X}$, then equality holds.

Proof: For the first part, define the events $A \triangleq \left\{i(\bar{\mathbf{X}};\mathbf{Y}) \geq i(\mathbf{X};\mathbf{Y})\right\}$ and $T \triangleq \left\{i(\mathbf{X};\mathbf{Y}) \geq t\right\}$ (T^c denotes the complementary event). Then:

$$\begin{aligned}
\epsilon_{\text{RCU}} &= \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot(1-Z)\right\}\right] && (23a)\\
&= \mathbb{P}\left(T^c\right)\cdot\mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot(1-Z)\right\}|T^c\right] \\
&\quad + \mathbb{P}\left(T\right)\cdot\mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot(1-Z)\right\}|T\right] && (23b)\\
&\leq \mathbb{P}\left(T^c\right) + \mathbb{P}\left(T\right)\cdot\mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[(M-1)\mathbb{P}\left(A|\mathbf{X},\mathbf{Y}\right)|T\right] && (23c)\\
&= \mathbb{P}\left(T^c\right) + (M-1)\mathbb{P}\left(T\right)\cdot\mathbb{P}\left(A|T\right) && (23d)\\
&= \epsilon_{\text{G-I}} && (23e)
\end{aligned}$$

For the second part, recall that by Lemma 1, $i(\bar{\mathbf{X}};\mathbf{Y})$ is statistically independent of $(\mathbf{X},\mathbf{Y})$. Denote by t^* the minimal threshold t such that

$$(M-1)\mathbb{P}\left(i(\bar{\mathbf{X}};\mathbf{Y}) \geq t\right) \leq 1.$$

Then $(M-1)\mathbb{P}\left(i(\bar{\mathbf{X}};\mathbf{Y}) \geq i(\mathbf{X};\mathbf{Y})\middle|i(\mathbf{X};\mathbf{Y}) < t^*\right) \geq 1$. Under the notation of U from the first part the proof, we have that: $\mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,U\right\}|i(\mathbf{X};\mathbf{Y}) < t^*\right] = 1$, i.e., the inequality in (23c) is equality in this case. ∎

Remark 2: It follows, that for the BSC, $\epsilon_{\text{G-I}} = \epsilon_{\text{RCU}}$. Indeed, it is noted in [1] that for the BSC, the RCU bound is equal to Poltyrev's bound [4]; this is not surprising, since the latter is derived from (21) (Poltyrev's bound uses linear codes, see Section V in the sequel).

Remark 3: Gallager's type I bound can be improved by breaking ties, similar to the improvement of RCU*, leading to G-I*. An analysis result to Proposition 3 relates G-I* and RCU*.

C. Threshold-Decoding Union Bounds

The average error probability of an RCSE can be further lower-bounded using the sub-optimal *threshold decoder* [5]. This decoder looks for a codeword that has a likelihood score above some predetermined threshold. In [1, Theorem 18] a union bound is derived for such a decoder, where if multiple codewords pass the threshold, the winner is chosen uniformly from among them.[5] The resulting "dependence testing" (DT) bound is given by:

$$\epsilon_{\text{DT}} \triangleq \mathbb{P}\left(i(\mathbf{X};\mathbf{Y}) \leq \gamma\right) + \frac{M-1}{2}\mathbb{P}\left(i(\bar{\mathbf{X}},\mathbf{Y}) > \gamma\right), \quad (24a)$$

where the optimal threshold is given by[6]

$$\gamma = \log\frac{M-1}{2}. \quad (24b)$$

A troubling behavior, demonstrated in [1] using the binary erasure channel (BEC), is that sometimes $\epsilon_{\text{RCU}} > \epsilon_{\text{DT}}$. This is counter-intuitive since the DT bound is derived from a sub-optimal decoder. We find that this artifact stems from the fact that the RCU bound ignores ties, and prove that the improved bound, denoted by RCU*, always satisfies $\epsilon_{\text{RCU}^*} \leq \epsilon_{\text{DT}}$. To that end, we prove a (very slightly) improved bound for the threshold decoder, that is closer in form to the ML bounds (13) and (21). It uses the following definitions (cf. (10)).

$$W_q = \mathbb{P}\left(q(i(\bar{\mathbf{X}};\mathbf{Y})) = q(i(\mathbf{X};\mathbf{Y}))|\mathbf{X},\mathbf{Y}\right) \quad (25a)$$

$$Z_q = \mathbb{P}\left(q(i(\bar{\mathbf{X}};\mathbf{Y})) < q(i(\mathbf{X};\mathbf{Y}))|\mathbf{X},\mathbf{Y}\right), \quad (25b)$$

where $q(i)$ is the indicator function:

$$q(i) = \mathbb{1}_{\{i>\gamma\}}. \quad (25c)$$

Proposition 4: For an RCSE,

$$\epsilon^{(\text{RCSE})} \leq \epsilon_{\text{TU}}, \quad (26)$$

where

$$\epsilon_{\text{TU}} \triangleq \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot\left(1-Z_q-\frac{W_q}{2}\right)\right\}\right]. \quad (27)$$

Furthermore, $\epsilon_{\text{TU}} \leq \epsilon_{\text{DT}}$.

Proof: For proving achievability, consider a decoder identical to the ML decoder, except that before comparing the words, the information-density scores are quantized according to (25c). For the comparison to the DT bound,

$$\begin{aligned}
\epsilon_{\text{DT}} &= E_{\mathbf{X},\mathbf{Y}}\left[\mathbb{1}_{\{i(\mathbf{X};\mathbf{Y})\leq\gamma\}} + \frac{M-1}{2}\mathbb{P}\left(i(\bar{\mathbf{X}},\mathbf{Y}) > \gamma\middle|\mathbf{X},\mathbf{Y}\right)\right] && (28a)\\
&\geq E_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,(M-1)\cdot\left[\mathbb{1}_{\{i(\mathbf{X};\mathbf{Y})\leq\gamma\}}\right.\right.\right. \\
&\quad \left.\left.\left. + \frac{1}{2}\mathbb{P}\left(i(\bar{\mathbf{X}},\mathbf{Y}) > \gamma\middle|\mathbf{X},\mathbf{Y}\right)\right]\right\}\right] && (28b)\\
&\geq E_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1,\frac{M-1}{2}\cdot\left[\mathbb{1}_{\{i(\mathbf{X};\mathbf{Y})\leq\gamma\}}\right.\right.\right. \\
&\quad \left.\left.\left. + \mathbb{P}\left(i(\bar{\mathbf{X}},\mathbf{Y}) > \gamma\middle|\mathbf{X},\mathbf{Y}\right)\right]\right\}\right] && (28c)\\
&= \epsilon_{\text{TU}}. && (28d)
\end{aligned}$$

[5] In fact, the proof states that the "first" codeword to pass the threshold is selected. However, such ordering of the codewords is not required.

[6] In [6], the threshold is further optimized, depending on the competing codeword and on the received word

■

Remark 4: It is not obvious that the optimal threshold for the TU bound is γ of (24b). However, it is good enough for our purposes.

Proposition 5: For any channel, $\epsilon_{\text{RCU}^*} \leq \epsilon_{\text{TU}}$. Thus, the RCU* bound is tighter than the DT bound, i.e.:

$$\epsilon_{\text{RCU}^*} \leq \epsilon_{\text{DT}}.$$

Proof: Recalling Remark 1, the RCU* bound reflects optimal (ML) pairwise decision. Thus, necessarily the pairwise error probabilities satisfy:

$$Z + \frac{W}{2} \geq Z_q + \frac{W_q}{2}. \tag{29}$$

■

Remark 5: In fact, the case of the BEC, where $\epsilon_{\text{RCU}^*} = \epsilon_{\text{TU}} = \min(1, \epsilon_{\text{DT}})$ is very special. In the BEC, an impostor cannot have a higher score than the true codeword; if the channel realization is such that the non-erased elements of $\mathbf{x}$ and $\bar{\mathbf{x}}$ are equal, then $i(\bar{\mathbf{x}};\mathbf{y}) = i(\mathbf{x};\mathbf{y})$, otherwise $i(\bar{\mathbf{x}};\mathbf{y}) = -\infty$. Thus,

$$\epsilon_{\text{RCU}^*} = \mathbb{E}_{\mathbf{X},\mathbf{Y}}\left[\min\left\{1, \frac{(M-1)W}{2}\right\}\right]. \tag{30}$$

Let k be the number of non-erased symbols out of the block of n, then $W = 2^{-k}$. Consequently, $(M-1)W/2 > 1$ if and only if $i(\mathbf{x},\mathbf{y}) < \gamma$, where γ is given by (24b).

D. Performance Comparison

Comparison of the different union bounds is given in Figure 2. In particular, the effect of tie-breaking on the bounds is shown by the comparison of the RCU bound (13) and the RCU* bound (14). Notice that this bound depends on the ensemble. Due to Lemma 1, the computation of the RCU and RCU* bounds for PIE becomes simple, hence show the bounds for this ensemble. Since an i.i.d. ensemble is also PIE, the exact error probability for i.i.d. ensemble (11) is given as a reference.

V. Linear Codes

Most known good codes are linear. Beyond that, linear codes have an important role for two reasons. First, they allow to improve performance (both capacity and error probability) in many network problems (see, e.g., [7], [8]). Second, for some channels, the average error probability and maximal error probability coincide for linear codes.

A. The Dithered Linear Codes Ensemble

For any finite field $\mathbb{F}_q$ of cardinality q, we define the dithered linear codes ensemble by

$$\mathbf{X}_j = H\mathbf{w}_j + \mathbf{D}. \tag{31}$$

Here, all operations are defined over the field, all elements of the $n \times k$ generator matrix H and length-k dither vector $\mathbf{D}$ are drawn uniformly and independently from the field elements, and $\{\mathbf{w}_j\}$ are all k-vectors over the field. It follows that the codebook size is $M = q^k$. An important special case is when r is prime, and modulo arithmetic is used, e.g., binary (dithered) linear codes.

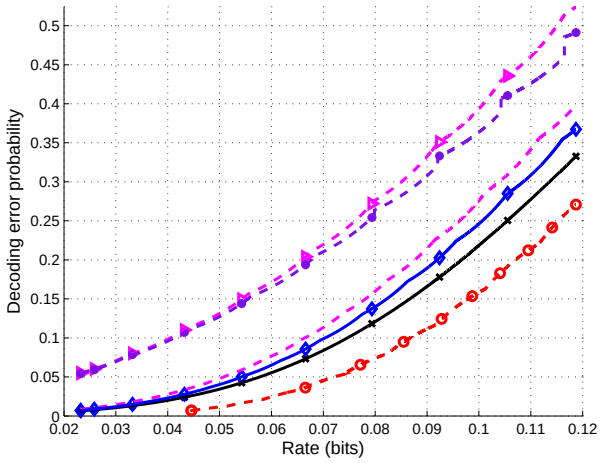

Fig. 2. The effect of tie-breaking on the performance of PIE codebooks of the different union-bounds. We demonstrate the effect using a BSC with crossover probability 0.3, at blocklength $n = 100$. The triangle-marked dashed curve is the DT bound (24). The asterisk-marked dashed curve is the TU bound (27). The dashed curve is the RCU bound (13). The diamond-marked solid curve is the RCU* bound (14). For reference, we repeat two of the curves of Figure 1. The $\times$-marked solid curve is the exact performance of the i.i.d. ensemble (11), while the circle-marked dashed curve is the lower bound for any codebook [1, Theorem 16]. The non-smoothness of some of the curves is not an artifact, but comes from the fact that they involve integers.

By [9], any dithered linear codes ensemble over this field is PIE. Consequently, the RCU* bound applies to this ensemble. Further, it is proven in [1, Appendix A] that for a class of channels, which includes the BSC and the BEC, there exists a *randomized* ML decoder such that the *maximal* error probability $\overline{\epsilon}$ (3) coincides with the average one.

B. Additive Channels

We now restrict our attention to channels that are additive, in the sense of Definition 3. Further, assume that the channels are additive over a finite field, which is the same field over which the code is linear. Clearly, in this situation the dither does not change the distance profile of the codebook, thus it suffices to consider the linear codes ensemble

$$\mathbf{X}_j = H\mathbf{w}_j, \tag{32}$$

where again H is i.i.d. uniform over $\mathbb{F}_q$. More importantly, in order to achieve good maximal error probability, there is no need to use randomized decoders.

Theorem 2: For any channel that is additive over a finite field, for an ensemble of linear codes over the field, there exists a deterministic decoder satisfying:

$$\overline{\epsilon} \leq \epsilon_{\text{RCU}^*}$$

Remark 6: Recall that the size of linear code is $M = q^k$ for an integer k. Thus, the theorem does not give $\bar{\epsilon}$ for all n, M.

Proof: Let $\Omega_1, \ldots, \Omega_M$ be a partition of the output space $\mathcal{Y}^n$ into decision regions (for any $1 \leq m \leq M$, Ω_m is associated with codeword m). A partition is optimal in the average error probability sense, if and only if it satisfies:

$$\Omega_m \subseteq \{\mathbf{y} \in \mathcal{Y}^n | \forall m' \neq m : V(\mathbf{y}|\mathbf{x}_m) \geq V(\mathbf{y}|\mathbf{x}_{m'})\} \quad (33a)$$

$$\Omega_m \supseteq \{\mathbf{y} \in \mathcal{Y}^n | \forall m' \neq m : V(\mathbf{y}|\mathbf{x}_m) > V(\mathbf{y}|\mathbf{x}_{m'})\}, \quad (33b)$$

and for all $m \neq m'$ $\Omega_m \cap \Omega_{m'} = \emptyset$. By (7), we have that for an additive channel, (33) is equivalent to

$$\Omega_m \subseteq \left\{\mathbf{y} \in \mathbb{F}_q^n \middle| \forall m' \neq m : P_{\mathbf{N}}(\mathbf{y} - \mathbf{x}_m) \geq P_{\mathbf{N}}(\mathbf{y} - \mathbf{x}_{m'})\right\}. \quad (34a)$$

$$\Omega_m \supseteq \left\{\mathbf{y} \in \mathbb{F}_q^n \middle| \forall m' \neq m : P_{\mathbf{N}}(\mathbf{y} - \mathbf{x}_m) > P_{\mathbf{N}}(\mathbf{y} - \mathbf{x}_{m'})\right\}. \quad (34b)$$

Since for any such optimal partition $\epsilon \leq \epsilon_{\text{RCU}^*}$, it is sufficient to show that there exists a partition satisfying (34) for which $\bar{\epsilon} = \epsilon$.

Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a linear code, and without loss of generality assume that $\mathbf{x}_1 = \mathbf{0}$. We construct decoding regions for $\mathcal{C}$ in the following way. Recall that $\mathcal{C}$ induces a (unique) partition of the linear space $\mathbb{F}_q^n$ into disjoint cosets $\mathcal{S}_1, \ldots, \mathcal{S}_{n-k}$, where each coset is a translation of the linear code. For each coset $\mathcal{S}_j$ let the coset leader $\mathbf{y}_{j,1}$ be some word that may belong to Ω_1 according to (34a). By definition, all coset elements of the coset can be uniquely labeled as

$$\mathbf{y}_{j,m} = \mathbf{y}_{j,1} + \mathbf{x}_m.$$

Assign each word $\mathbf{y}_{j,m}$ to Ω_m. It then satisfies that for all $m' = 1, \ldots, M$

$$\begin{aligned} P_{\mathbf{N}}(\mathbf{y}_{j,m} - \mathbf{x}_m) &= P_{\mathbf{N}}(\mathbf{y}_{j,1}) && (35a)\\ &\geq P_{\mathbf{N}}(\mathbf{y}_{j,1} - (\mathbf{x}_{m'} - \mathbf{x}_m)) && (35b)\\ &= P_{\mathbf{N}}(\mathbf{y}_{j,m} - \mathbf{x}_{m'}), && (35c)\end{aligned}$$

where in (35b) we have used the facts that $\mathbf{y}_{j,1}$ satisfies (34a) and that the sum of codewords is a codeword. It follows, that the partition indeed satisfies (34). To see that $\bar{\epsilon} = \epsilon$, we have that for all $m = 1, \ldots, M$:

$$\begin{aligned} \mathbb{P}(\mathbf{Y} \in \Omega_m | \mathbf{X} = \mathbf{x}_m) &= \mathbb{P}(\mathbf{X} + \mathbf{N} \in \Omega_m | \mathbf{X} = \mathbf{x}_m) && (36a)\\ &= \mathbb{P}(\mathbf{N} \in \Omega_m - \mathbf{x}_m | \mathbf{X} = \mathbf{x}_m) && (36b)\\ &= \mathbb{P}(\mathbf{N} \in \Omega_1 | \mathbf{X} = \mathbf{x}_m) && (36c)\\ &= \mathbb{P}(\mathbf{N} \in \Omega_1) && (36d)\\ &= \mathbb{P}(\mathbf{x}_1 + \mathbf{N} \in \Omega_1) && (36e)\\ &= \mathbb{P}(\mathbf{Y} \in \Omega_1 | \mathbf{X} = \mathbf{x}_1), && (36f)\end{aligned}$$

where (36c) is due to the construction of Ω_m, and (36d) is since the noise is statistically independent of the channel input. ■

Remark 7: The partition used in the proof is not unique, in the sense that for some cosets the choice of the coset leader is arbitrary. However, for any such choice the coset is partitioned in a fair way between the decision regions.

REFERENCES

[1] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[2] G. Seguin, "Linear ensembles of codes," *IEEE Trans. Information Theory*, vol. 25, no. 4, pp. 477 – 480, July 1979.

[3] R. G. Gallager, "Low-density parity-check codes," Ph.D. dissertation, Massachusetts Institute of Technology, 1963.

[4] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. Information Theory*, vol. 40, no. 4, pp. 1284 –1292, July 1994.

[5] A. Feinstein, "A new basic theorem of information theory," *Information Theory, Transactions of the IRE Professional Group on*, vol. 4, no. 4, pp. 2 –22, September 1954.

[6] A. Martinez and A. Guillén i Fàbregas, "Random-coding bounds for threshold decoders: Error exponent and saddlepoint approximation," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, Aug. 2011.

[7] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Information Theory*, vol. IT-57, pp. 6463–6486, Oct. 2011.

[8] E. Haim, Y. Kochman, and U. Erez, "Distributed structure: Joint expurgation for the multiple-access channel," *IEEE Trans. Information Theory*, submitted, 2012. online available: http://arxiv.org/abs/1207.1345.

[9] R. L. Dobrushin, "Asymptotic optimality of group and systematic codes for some channels," *Theor. Probab. Appl.*, vol. 8, pp. 52–66, 1963.